MobiHoc Poster: Power Efficient Gathering of Correlated Data: Optimization, NP-Completeness and Heuristics*

Răzvan Cristescu
Razvan.Cristescu@epfl.ch

Martin Vetterli
Martin.Vetterli@epfl.ch

Swiss Federal Institute of Technology (EPFL), Lausanne CH-1015, Switzerland

This paper studies the interaction between the communication costs in a sensor network and the structure of the data that it measures. We formulate an optimization problem for power efficient data gathering and show that the problem is NP-complete. We propose scalable, distributed and efficient heuristics for solving this problem and show by numerical simulations that the power consumption can be significantly improved over direct transmission or the shortest path tree. Our algorithms provide solutions close to a computationally heavy heuristic used as benchmark, simulated annealing, which is provably optimal in the limit.

I. Correlated data gathering

Sensor networks measure data which is usually not independent at different locations, but rather correlated. In *data gathering*, there is one sink node (the base station), and all other nodes are information sources. All their data need to arrive at the base station. This problem has been addressed in [2], [3], [4], [5]. Our novel approach is to exploit the interaction of source coding and transmission, for improving power efÞciency of correlated data gathering. In our setting, the data structure (correlation) inßuences the communication structure (the tree built for gathering the data).

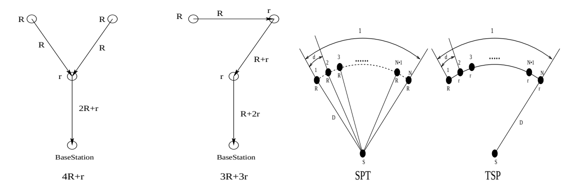
II. System model and optimization

As battery power is the scarce resource, the total power used by the network has to be minimized. The power needed to transmit data from a node essentially consists of the product between the aggregated amount of data transmitted by that node, and the weight of the link to its parent node in the gathering tree. The weights on the links are functions of the distances between the nodes.

We consider a simpliÞed model of interaction between data supply at nodes and transmission structure; however, our model preserves the original complexity of the problem. We will show that the combined treatment of both source coding and transmission makes the problem NP-complete. The reason is that the data amounts supplied at nodes *depend* on the transmission structure, due to correlation in the data. On the contrary, in classical network transport theory, supplies at nodes are *fixed*.

Denote by X_i the random variable measured at node i, and by $d_{i,j}$ the link weights. Data at nodes without side information are coded with $H(X_i) = R$ bits. However, intermediate nodes on gathering paths do have side information available from their children. We assume that the reduction in entropy is independent from the distance and the amount of side information available: $H(X_i|X_j, \dots) = H(X_i|X_j) = r$ bits, $j \neq i$, and $0 \leq r \leq R$. Let

*The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communications Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.



(a) All edge weights are 1. (b) *SPT* vs. *TSP*

Figure 1: Simple network examples.

$\rho = 1 - r/R, 0 \leq \rho \leq 1$ be the correlation coefÞcient. When ρ is 1, the data are strongly correlated; when ρ is 0, the data are independent.

Finding good correlated data gathering trees is not trivial even for very simple networks (Figure 1). If data were independent, the shortest path tree (*SPT*) would be optimal. However, in Figure 1(a), as soon as $\rho > 1/2$, the *SPT* is no longer optimal. In Figure 1(b), the ratio of total used powers is $\lim_{N\to\infty} \frac{p_{TSP}}{p_{SPT}} = (1-\rho)\left(\frac{1}{2D}+1\right)$. If $\rho = 1$, a traveling salesman path (*TSP*) is arbitrarily more power efÞcient than direct transmission (*SPT*).

In general, the optimization problem is to Þnd the spanning tree ST that minimizes:

$$\rho \sum_{l \in L} d_{ST}(l, S) + (1-\rho) \sum_{i \in V} d_{ST}(i, S), \qquad (1)$$

where V is the set of nodes in the network, $L \subset V$ is the set of leaves of ST, and $d_{ST}(i, S)$ is the total weight of the path connecting i to S on the ST tree.

If $\rho = 0$, the optimal tree is *SPT* (polynomial time). If $\rho = 1$, the optimal solution is the *multiple* TSP (NP-complete). In section III we show that the problem is NP-complete in the general case $0 < \rho \leq 1$ as well. We present in section IV heuristics that provide good gathering trees.

III. NP-completeness

Definition 1. MINIMUM POWER GATHERING TREE

INSTANCE: A undirected weighted graph $G = (V, E)$, a node $S \in V$, a positive integer M.

QUESTION: Does the graph admit a spanning tree ST of cost, as deÞned in (1), at most M?

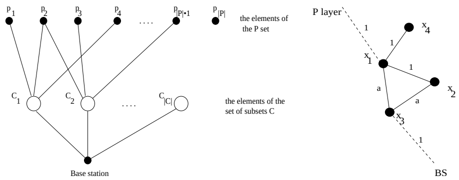


Figure 2: (a) A graph instance; (b) Gadget for C_i.

THEOREM 1. *There is no polynomial time algorithm that solves the MINIMUM POWER GATHERING TREE problem, unless P=NP.*

PROOF. Our proof is based on a reduction from the MINIMUM COVER problem [1]. An instance of the MINIMUM COVER problem is a collection C of subsets of a Þnite set P, and a positive integer $K \leq |C|$. For each such instance, we consider a three layers graph instance of our problem (Figure 2(a)). For each $C_i \in C$ we build a structure as in Figure 2(b). Let $M = |P|(d+a+1)R + K(2aR+3R+a+2) + (|C|-K)(aR+3R+2a+4)$. Finding a spanning tree with cost at most M is equivalent to Þnding a set cover for the set P, of cardinality at most K. The construction of the graph instance is polynomial, so our problem is at least as hard as MINIMUM COVER, and thus NP-complete. $\square$

IV. Algorithms and simulations

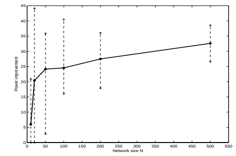
Leaves deletion heuristic

We start with *SPT* as initial guess. Nodes maintain only local information: parent, number of children, $d_{ST}(i,S)$. Then, as long as power improvements are obtained, every leaf node i Þnds in its neighborhood the leaf node j that maximizes $R(d_{ST}(i,S) + d_{ST}(j,S)) - (R(d_{i,j} + d_{ST}(j,S)) + rd_{ST}(j,S)) - I(i)$. If this quantity is positive, $par(i) \rightarrow j$, and all necessary updates are done for i, former $par(i)$, and j. This algorithm involves only $3-4$ supplementary steps after *SPT* is computed, and is fully distributed.

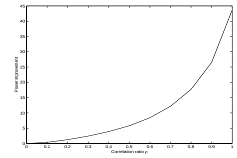
Balanced SPT / multiple TSP tree

This heuristic is a combination of the *SPT* and multiple *TSP*. We Þrst build the *SPT* for nodes within a radius $q(\rho)$ away from the base station (Figure 4(c)). Then, successively add to the tree node i that minimizes $d(i,l) + d_{ST}(l,S)$, where $l \in L$ are leaves of the current subtree. This is a simple suboptimal nearest neighbor approximation of the multiple *TSP*.

Our simulations were done in MATLAB for a network of up to $N = 500$ nodes randomly distributed on a square grid. The *SPT* was found with the distributed Bellman-Ford algorithm, that runs in $\mathcal{O}(N|E|)$ steps. Our extensive experiments show important improvements (up to 40 $\%$) of the leaves deletion and the balanced *SPT/TSP* heuristics over *SPT*, for randomly distributed nodes (Figure 3, 4).

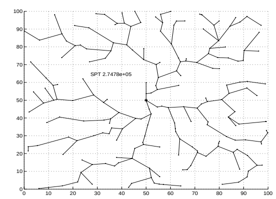


(a) Power gain vs. N.

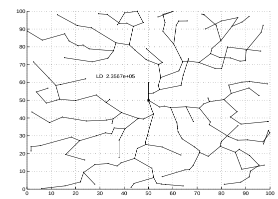


(b) Power gain vs. ρ.

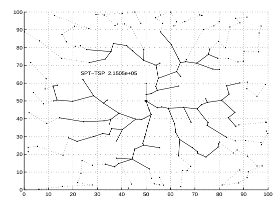
Figure 3: Average power improvement (in %) of leaves deletion (LD) over shortest path tree (*SPT*) for (a) $\rho = 0.9$, and (b) $N = 200$.



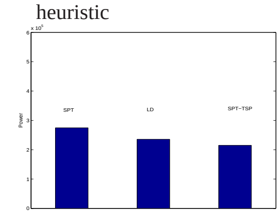
(a) *SPT*



(b) Leaves deletion heuristic



(c) Balanced *SPT-TSP*



(d) Power efÞciency

Figure 4: Data gathering tree on a random network instance. $N = 200$, $\rho = 0.2$.

References

[1] M.R. Garey and D.S Johnson. *Computers and intractability*. W.H. Freeman, 1979.

[2] A. Goel and D. Estrin. Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk. In *ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[3] W. Rabiner Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efÞcient communication protocol for wireless microsensor networks. In *Proc. of HICSS 2000*.

[4] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proc. of ACM MOBICOM 2000*.

[5] S. Lindsey, C. S. Raghavendra, and K. Sivalingam. Data gathering in sensor networks using the energy*delay metric. In *Proc. of IPDPS Workshop on Issues in Wireless Networks and Mobile Computing*, 2001.